A Computerized Phonetics Instructor: BABEL

Joaquin Vila
Illinois State University

Lon Pearson
University of Missouri—Rolla

ABSTRACT: Babel is an expert system able to animate (graphically) and reproduce (acoustically) a text in any language which uses the Latin alphabet. This system has been developed to aid language learners and to help instructors leach the fine nuances of phonemes. Each phoneme has a unique sound and thus requires a precise positioning of the vocal organs which are displayed on the screen in two different projections: a front view and a profile cross view of a human face in synchronization with the output sounds of the speech synthesizer.

KEYWORDS: CALL, expert systems, computer graphics animation, phonetics, speech synthesizer, text-to-speech.

In the hands of teachers and students alike, the Babel language teaching system is an innovative and exciting tool. It has taken advantage of recent developments in computer graphics, speech synthesis, and artificial intelligence to produce a computerized visual and auditory speech model. Teachers can use Babel as an audio-visual and auditory speech model. Teachers can use Babel as an audio-visual aid, and students can use it as a tutorial system to help them learn correct positioning of speech organs.

Babel can be used in the field of education by teachers of Foreign Languages (FL) and English as a Second Language (ESL). Speech pathologists working with children or teaching the hearing impaired will also find it invaluable.

Equally important, Babel is interactive. Students in any of these areas can easily learn how to make Babel speak to them, which will allow them to visualize the way to form speech, showing just how and where certain sounds and speech patterns are pronounced. This comprehension of where to put the tongue or lips, or how wide to open the mouth, is indispensable for the formation of correct speech.

As every teacher of language knows, one of the most critical difficulties that students encounter in learning a foreign language is understanding how to pronounce properly unfamiliar sounds demanded by the FL. Students, especially older ones, have become so accustomed to using only the sounds required by

their mother tongue, that they often cannot conceive how other sounds are produced. Thus when they attempt the pronunciation of new sounds demanded by other languages, they have trouble both in conceptualizing such new and rare sounds and in producing them. Apart from ingrained habits that are hard for individuals to break, a great amount of the novices difficulty stems from their inability to hear accurately the new sounds of the target language and to be able to discriminate subtle sound differences (phonemes and allophones). They can neither attain the fine tuning required nor see inside the mouth to distinguish the sounds. The unfortunate result is that many students still cannot pronounce such sounds, even after repeated classroom drill. This becomes a critical problem for both the teachers and the students. But with Babel as a teaching aid, the viewer can see the correct places of articulation on the computer screen and can hear words and sentences pronounced correctly by a speech synthesizer.

On the screen, Babel displays two animated projections of the human face: the first graphic is a front view of a face and the second is a traditional phonetician's cutaway side view of the throat and jaw. At the bottom of the screen on a text line, the user types in words to be pronounced. In response to the user's keyboard input Babel also reproduces acoustically the text typed onto the screen. In other words, Babel reacts to the user input by speaking those words typed, and by displaying in screen windows both frontally and laterally (by showing moving lips as well as cross-sectioned speech organs) just how that sound is correctly produced.

The first part of this article presents Babel in general terms; then for those who would like more information regarding artificial intelligence and how Babel operates, the last part of the article will discuss programming concepts and will describe Babel's components: a rule-editor and a rule-interpreter.

Babel began as a graduate computer project and masters thesis in the Institute of Artificial intelligence at the University of Missouri. Spanish is the Natural Language we selected as a model for all the examples and illustrations, because Spanish presented a clear-cut, workable phonology. Also the authors have a solid background in Spanish. Moreover, Babel is also adaptable to English, German, French, and other western languages which use the Latin alphabet. Only the "knowledge base" of the new FL has to be developed using the rule editor to allow the expert system to make a successful phonetic transcription of the new target language.

Human phonetics is complicated but limited at the same time. The number of sounds which human beings are potentially capable to emit using their speech organs is immense. However, each language has a unique pattern of sounds. Tomas Navarro Tomas asserts that: "Some phonemes are of universal extent; others are found only in certain languages. Phonemes of a general character do not appear in the same proportion in all languages. The sound image of a language depends greatly on the proportion it uses the phonemes

with [sic] and specially on the particular modality it follows within the number of variants that such units permit. In describing the oral shapes of the word, it is difficult to establish precise boundaries between sound and phoneme, between phonetics and phonology. At any rate, the general appearance of sounds, the effects produced by their combinations, and, especially, the role they play in relation to the meaning of words are all part of phonology" (1968, 14).

The Spanish official orthography, though more phonetic than other languages, is not even close to an appropriate representation of its pronunciation. The phonological series of Spanish consists of forty-two phonemes. The number of variants (allophones) that these phonemes assume in the pronunciation of all the countries where this language is spoken is incalculable.

However, knowledge of the frequency of the phonemes in each language was relevant to develop the knowledge base of the Babel expert system. Tomas Navarro Tomas stated that "the rate of frequency of phonemes is an indispensable norm for knowing the composition of each language, for comparing languages, and for indicating the appropriate order in the teaching of pronunciation" (1968, p. 17). It is important that one be careful in generating the rules of pronunciation of the phonemes with high frequency. Navarro Tomas asserts that the vowels a, e, o, and the consonant s represent 40% of the phonetic material used in any Spanish written text. A second category is that formed by n, r, 1, d, t, i. A third category belongs to k (c, q) m, p, b, z, u, and g. And finally the phonemes with less frequency are: rr, f, j, 11, y, n, ch, and the diphthongs and triphthongs of the language. Table I reproduces the proportions established by Navarro Tomas (p. 25-26).

TABLE I
Frequency of Spanish Phonemes

Vowels		Diphthongs	
a	13.00%	ie	0.86%
e	11.75	ia	0.54
o	8.90	ue	0.52
i	4.76	io	0.32
u	1.92	ua	0.20
		ai	0.15
		ei	0.15
		oi	0.15
	40.33%	au	0.09
		eu	0.05
		iu	0.05
		ui	0.05
		uo	0.03
		ou	0.00

		3.16%
Voiced Consonants	voiceless	consonants
n 6.94%	s	7.50%
r 5.91	t	4.82
l 5.46	k	4.23
d 5.00	p	3.06
m 3.09	z	2.23
b 2.54	f	0.72
g 1.04	j	0.51
s 1.00	ch	0.30
rr 0.80		
ll 0.60		23.37%
y 0.40		
n 0.36		
33.14%		

We chose for the student visual-training model (one of the screen images) a cross-sectioned diagram of the speech organs, because it is the most commonly accepted method of showing positions of speech (points of articulation). But to enhance the side view—to give the user a more life-like, natural image—a frontal view of a face that talks also appears in a window. In this front-view visual aid, graphically animated lips are superimposed on the face of a beautiful woman (see figure 1). The front-face window offers a more holistic view of the speech process, and is of special interest for speech pathologists and those who work with the hearing impaired.

These graphic images produce a lasting impression, and they effect for the student a valuable source for insights as to how lips, tongue, and mouth produce speech sounds.

Artificial Intelligence and Expert Systems

Our own problems in learning and teaching languages made us aware of the need to develop a computerized teaching device. We then set out to research both linguistics and artificial intelligence to discover how computers might be used to solve language learners' problems.

Our research on FL pronunciation problems showed that a solution for teaching purposes could be achieved using an expert system. An expert system is a sophisticated computer program that solves complicated problems using an accumulated knowledge base that has been gleaned from the wisdom of a

Figure 1. Babel's projection of the human face

human being who is an expert in that particular field. Expert systems present a favorable framework for phonetic transcription because they allow us to generate text-to-speech rules easily; and, moreover, these rules can then be updated without great effort. When developing the text-to-speech rules, the rule sequence and determination are not evident. So extensive modifications to the rules are necessary. Because conventional computer systems combine data and logic in the program, it is difficult to modify them. An expert system, however, allows users to modify the program smoothly due to its architecture.

We wanted a multi-language tool that could update the pronunciation rules (intelligence) demanded by the language being dealt with at the moment. The system that we devised carried out the necessary phonetic digital screening process by using a rule-interpreter (inference system). The rule-interpreter is a sieve-like algorithmic program that strains and selects through a computer-code matching process the rules to be applied. Next it cues the system which then begins to actually utilize the letter-to-sound rules to any input text. Once we were able to see that the structure of the design works, the next step was to translate text-to-speech auditory signals and synchronize them with the graphic animation images of the two projections of the human face in the previously mentioned windows. We wanted user-friendly software to make it effective and hardware that was affordable to users. The prototype was named Babel after the Biblical profusion of tongues.

The Babel system was designed to run on an IBM Personal Computer (or MS-DOS "compatible") with graphics capabilities. The PC also needs to be equipped with a Votalker IB, which embodies the Votrax SC-02 phoneme synthesizer. The Votalker IB incorporates 64 standard phonemes with the additional capability of producing allophones (variations upon phonemes). We realize the linguistic limitations of such inexpensive equipment.

The speech synthesis model began with the sound spectrograph invented during World War II. A marriage between digital electronics and linguistics, the spectrograph displayed in voiceprints details of uttered vocal patterns by showing sound waves of voice timbres.

Later several text-to-speech systems were developed, as well as other approaches which were created (some embodying large pronunciation dictionaries or linguistic analysis) although many were not practical. One model for Babel was the successful text-to-speech system by rule developed by the Naval Research Laboratory (NRL). Details about the system were published in December, 1976 under the title "Letter-to-Sound rules for Automatic Translation of English Text to Phonetics" (Elovitz [1976]). The NRL system demonstrated the practicality of routine text-to-speech translation. A set of 329 letter-to-sound rules was developed. Actually these rules translate English text into the international phonetic alphabet (IPA), producing correct pronunciations for approximately 90 percent of the words. A second set of rules translates IPA into the phonetic

coding for a particular commercial speech synthesizer.

The rule structure developed by the NRL team is analogous to that used in creating Babel. However, variations were made to the NRL system in order to generate a flexible rule syntax for Babel that would be capable of including the requirements of diverse languages.

Early in the 1980s Texas Instruments Inc. (TI) developed a powerful text-to-speech system with notable features (Fisher [1983]). However the source rule format of TI's system is a quasi-linguistic generalization of that used by the NRL system. Among all its features the most significant contribution TI made was the introduction of User Defined Symbols (UDS), which will be explained later.

Another facet of innovative graphic devices is the branch of articulatory synthesis. The first articulatory synthesis system was proposed and developed by Coker and Fujimura at the end of the 1960s. A method for generating synthetic speech was devised by them in which synthesizer control signals are derived by rules from phonetic input data through intermediate-step vocal-tract area computations. Phonemes, the basic elements of the input data, are characterized as static, context-independent, ideal, vocal-tract shapes. These are tabulated in the Coker-Fujimura program as sets of parameters for the vocal tract model. This proposal was developed later on a Honeywell DDP-516. However, by current standards it now appears to be an unrefined system.

A Lip-Reader Trainer system was written by Robin L. Hight of St. Louis. This software package converts typed input sentences into a corresponding sequence of lip, teeth and tongue positions on a graphics display (for an Apple 11). The system, which was intended to aid deaf people, only shows the positions of the lips in animation when a text is input to the system in phonetic form. The lip-reader trainer's contribution to Babel's existence is the knowledge that there is only a limited set of lip positions distinct enough to be read clearly by humans. With only nineteen possibilities in English, lip positions are sufficiently unambiguous to the users so that on phoneme can be distinguished from another. Of course, other FLs have some strikingly visible variances, such as the French u.

BABELS'S Architecture

The components of Babel are a rule-editor and a parser-like rule-interpreter (inference system).

A. Rule-Editor

The Rule-editor is the core of the system. With it, letter-to-sound rules can be developed to translate text to speech. The Rule-editor was provided with a very friendly interface to create and update the pronunciation rules of different languages. One accesses the rule-editor only to create and update knowledge.

The Rule-editor is mainly composed of four windows. At the left side of the screen is the WORKING RULE AREA where the rules are defined; in the

middle is the MENU AREA where the main menu and edit menu are displayed; at the right is the INFORMATION AREA where the User Defined Symbols are exhibited, the phoneme chart, and the character chart; and finally at the bottom is the INPUT/OUTPUT AREA where information relevant to the knowledge to load or save is supplied as well as all the operations involving User Defined Symbols and other utilities. Figure 2 and Figure 3 show two different states of the Rule-editor where all the windows can be recognized.

1. Rule Syntax: The rule formalism of this system is very similar to that of the NRL system. However variations were made in order to increase the rule's possibilities.

Each rule has the form:

A[B]C=D

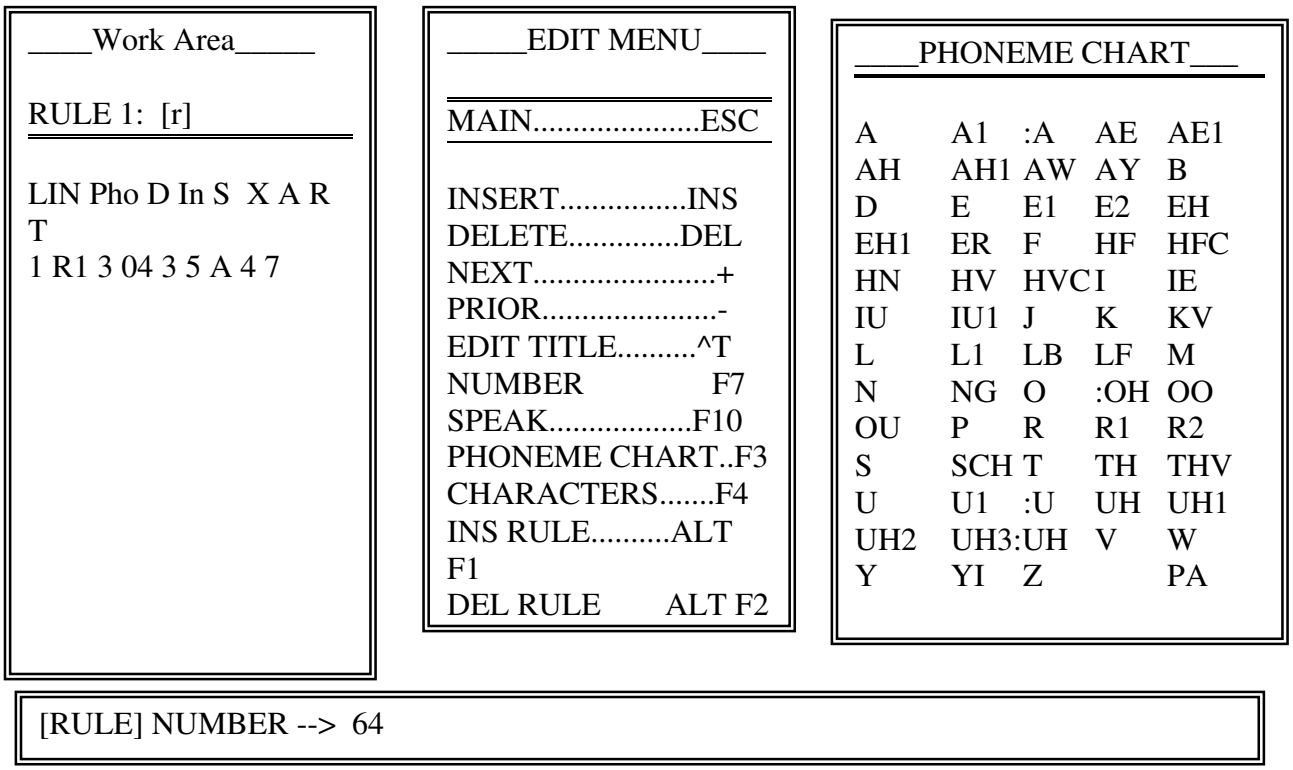


Figure 2. Rule-Editor (Edit Menu)

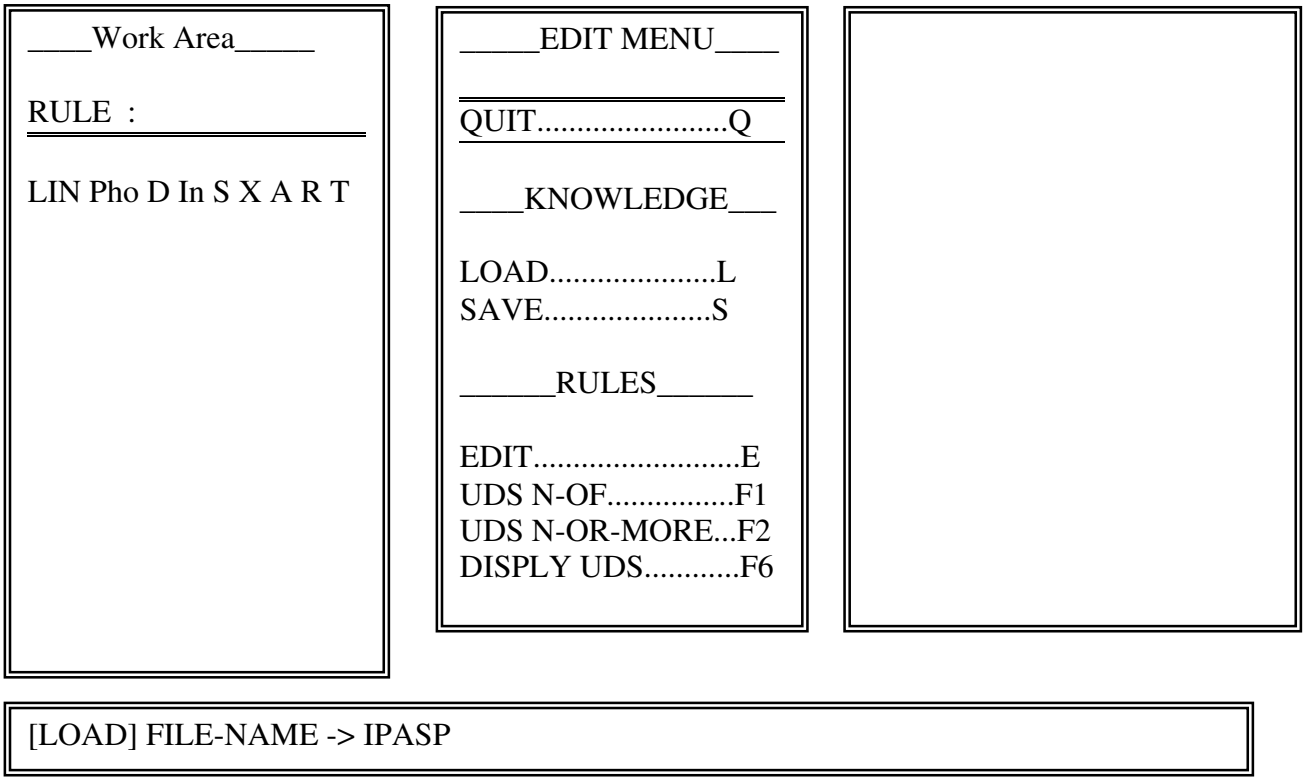


Figure 3: Rule-Editor (Main Menu)

The character string B (body rule), occurring with left context A (prefix rule) and right context C (suffix rule), induces the pronunciation D (rule consequence or value).

- D is one or more phonemes, or, in other words, is one or more of the 64 Votrax input symbols. See Table II. Each of these phonemes can be altered through the rule editor to produce allophones by adjusting one of the five speech parameters provided by the Votrax SC-02 synthesizer: duration, inflection, slope, pitch extension, and filter frequency.

TABLE II

Votrax phonemes					
Symbol	Votrax	Example	Symbol	Votrax	Example
[]	PA	(pause)	[1]	L	lady
[i]	E	keep, eat	[1]	L1	Louvre
[ ]	E1	become	[1]	LF	call
[e]	Y	marry	[w]	W	want, why
[]	YI	year	[b]	B	big
[a]	AY	made	[d]	D	said

[ ]	IE	ear	[g]	KV(HVC)	give
[]	I	mit	[p]	P	part
[e]	A	made	[t]	T	taste
[e]	A1	attainment	[k]	K	kite
[E]	EH	said	[*]	HV	(voiced)
[E]	EH1	enter	[g]	HVC	(g)
[oe]	AE	can	[h]	HF	hand
[oe]	AE1	happy	[*]	HFCT	(k)
[a]	AH	pop	[*]	HN	(m, n, ng)
[a]	AH1	honest	[z]	Z	zip, pays
[o]	AW	lost	[s]	S	sing, city
[o]	O	for	[3]	J	measure
[o]	OU	told	[S]	SCH	ship
[ ]	OO	look	[v]	V	vault
[ ]	IU	you	[f]	F	fat, phone
[ ]	IU1	should	[e]	THV	the, phone
[u]	U	you	[e]	TH	the, lathe
[ ]	U1	unit	[m]	M	man
[e]	UH	under	[n]	N	name
[e]	UH1	common	[n]	NG	long
[e]	UH2	constant	[*]	:A	Marchen
[e]	UH3	what	[*]	:OH	Lowe
[ ]	ER	word	[*]	:U	fun
[r]	R	ring	[*]	:UH	bluhen
[r]	R1	error	[*]	E2	bitte
[r]	R2	Mutter	[*]	LB	bluhen
*Unassigned					

- B is the character or character string to be translated. In this case, B can include all the Spanish letters with all the special characters, accents and exceptions. Figure 4 and Figure 5 display the accents and special letters (used in on-English languages) available in the Babel system, and how one can invoke them.

- A and C are the characters, string or special symbols (UDS user defined symbols) representing a class of character strings which denotes categories of sound such as vowels, voiced consonants, etc.

- Blanks are significant as they denote beginnings and ends of words.

- Rule-order is extremely important

- The absence of A or C in a rule means that the corresponding context is irrelevant.

The main difference between Babel and the NRL rule composition is that in Babel the expert is welcome to define his own special symbols (UDS), unlike the NRL system where symbols are already defined and do not facilitate the creation of rules for other languages.

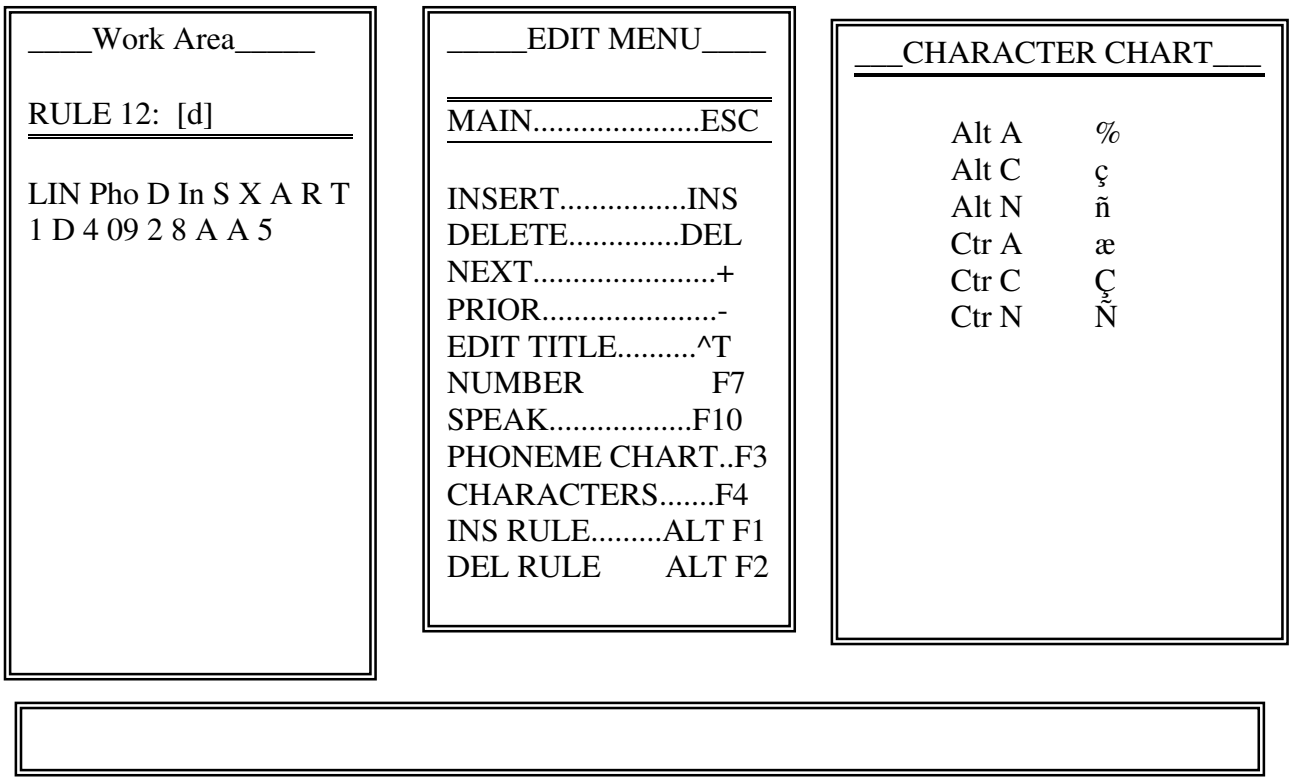


Figure 4. Special letters available in Babel

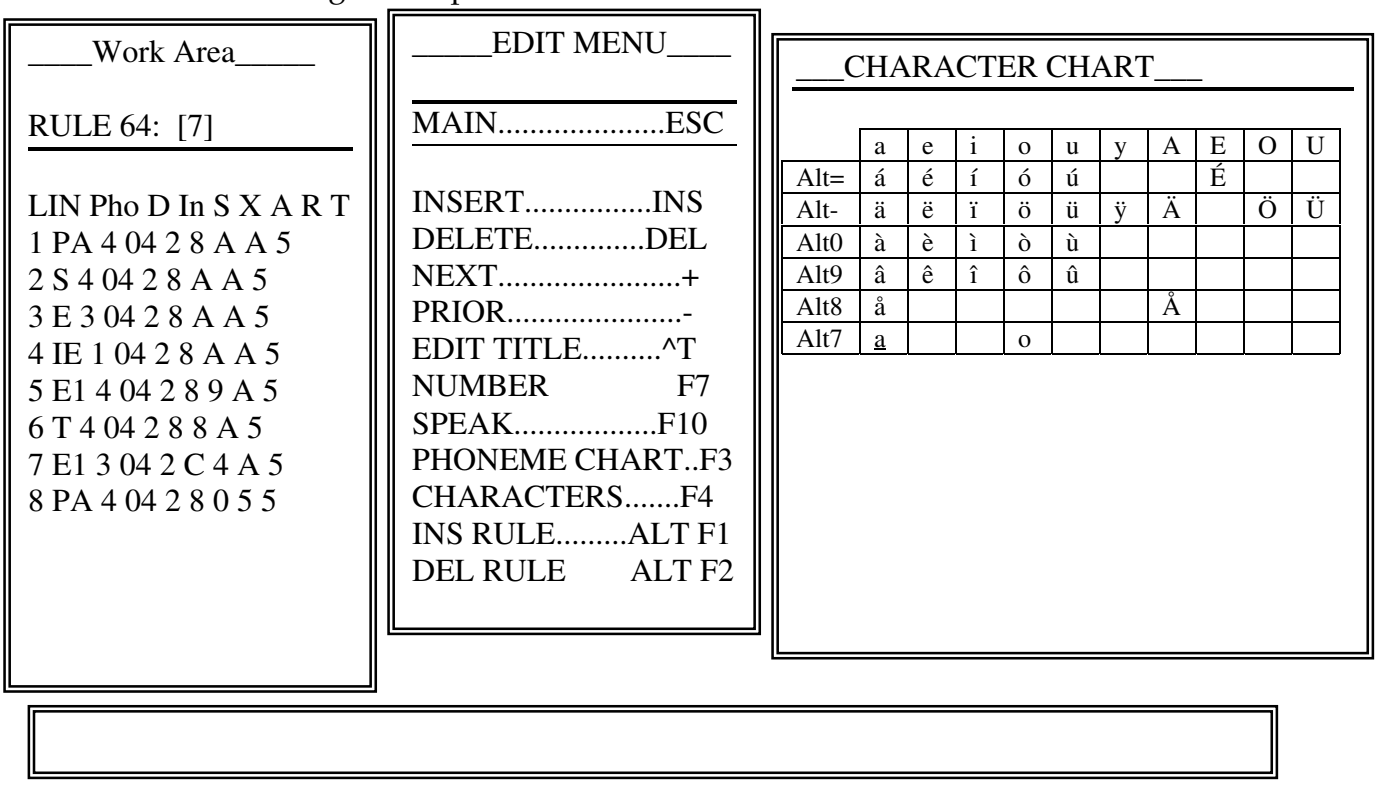

____Work Area_____
RULE 64: [7]
LIN Pho D In S X A R T
1 PA 4 04 2 8 A A 5
2 S 4 04 2 8 A A 5
3 E 3 04 2 8 A A 5
4 IE 1 04 2 8 A A 5
5 E1 4 04 2 8 9 A 5
6 T 4 04 2 8 8 A 5
7 E1 3 04 2 C 4 A 5
8 PA 4 04 2 8 0 5 5
_____EDIT MENU____
MAIN.....ESC
INSERT.....INS
DELETE.....DEL
NEXT.....+
PRIOR.....-
EDIT TITLE.....^T
NUMBER F7
SPEAK.....F10
PHONEME CHART..F3
CHARACTERS.....F4
INS RULE.....ALT F1
DEL RULE ALT F2
___CHARACTER CHART___
a e i o u y A E O U
Alt= á é í ó ú É
Alt- ä ë ï ö ü ÿ Ä Ö Ü
Alt0 à è ì ò ù
Alt9 â ê î ô û
Alt8 å Å
Alt7 a o

Figure 5: Special accents available in Babel

2. UDS (User Defined Symbols): The UDSs are special defined symbols representing a class of character strings which denote categories of sounds such as vowels, consonants, etc. The UDSs were introduced by Fisher [1983] in a text-to-speech development system. However, there are some variations in the process of defining a UDS in this system.

Babel supports two types of UDSS.

SYMBOL = n OR-MORE = (SET)
SYMBOL = n OR-MORE = (SET)

Where SYMBOL (#,$,%,&,*,+,A,: @) equals the number (n) of times an element of the set might appear; and (SET) is a list of character strings separated by commas. Examples of UDS are:

= 1 OR-MORE = A,E,I,O,U,Y
: = 0 OR MORE=B,C,D,F,G,H,J,K,L,M,N,O,P,Q,R,S,T,V,W,X,Z,
* = 1 OF =B,D,V,G,J,L,M,N,R,W,Z

Figure 6 displays the main menu of the Rule-editor and shows the process of defining a UDS.

A representative rule for English using a UDS (according to the previous UDS's examples) is

#:[e]

which means that an e at the end of a word preceded by # (one or more vowels) and : (zero or more consonants) is silent.

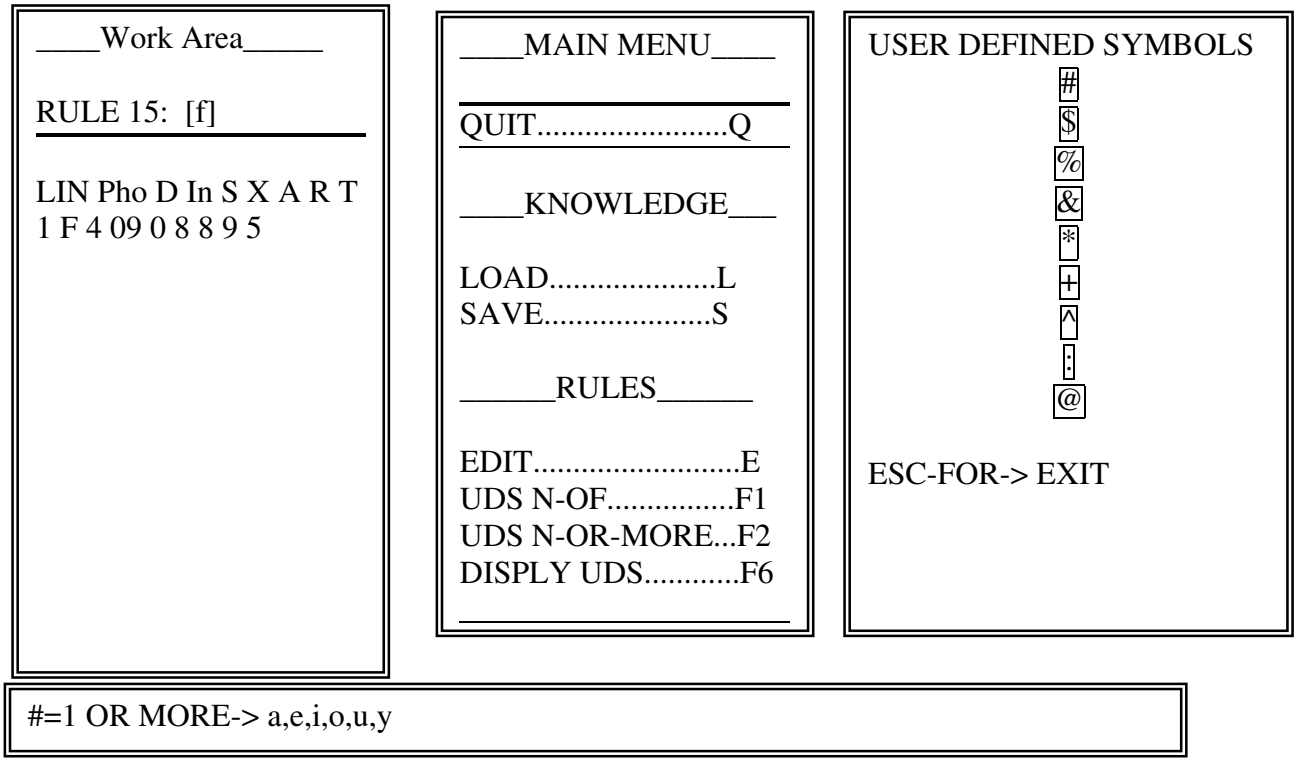


Figure 6. Rule-Editor (Edit Menu). The process of defining a UDS

B. RULE-Interpreter

The rule-interpreter is the pragmatic side of Babel. This subsystem has been designed to animate a human speech model using the stimuli of the input text and screening it through a set of rules which is loaded (knowledge selected) in the expert system.

The general block diagram exhibited in Figure 7 shows the process of the rule-interpreter, which involves the following:

1.Knowledge Selection.

- A welcoming display appears on the screen, requesting that the user choose the knowledge to be loaded. See Figure 8. (The rule-interpreter accepts any knowledge created by the rule-editor).

2. Load Rules.

- The rules bearing the name of the knowledge selected are loaded into the system.
- Next, two projections of the human face (front and profile cross view) are displayed on the screen. See Figure 1.

3. Input Text.

- The user is free to type any text. (The input text is echoed at the bottom of the screen, in the input window).

4. Phonetic Transcription.

-The expert system scans the text and produces a phonetic transcription of it.

- The phonetic transcription process is: "The process of transcribing a spoken word [text] into its phonetic components..." (Votalker IB 1985).

-The phonetic transcription process involves the following

*The input text is scanned from left to right.

*Then the subset of rules pertinent to the single character pointed to at any given time is scanned.

*The rule-interpreter decodes and applies the rules until a rule triggers.

*The value 'D' of the rule triggered (the sequence of phonemes) is then transmitted to a temporary buffer. *The last rule in the scanned subset is always the default pronunciation of 'B' (body rule or character string to be translated).

*The pointer advances as many characters over the source text as the number of characters of 'B' (the body rule). *The scan process is over when all the characters of the source text are exhausted.

-Table III shows how the phrase "le rogue, Enrique" is scanned.

5.Animation of Speech (Image and Sound).

-A succession of pictures showing the vocal speech organs of each phoneme generated by the phonetic transcription are exhibited on the screen at the same time that the sounds are uttered by the synthesizer.

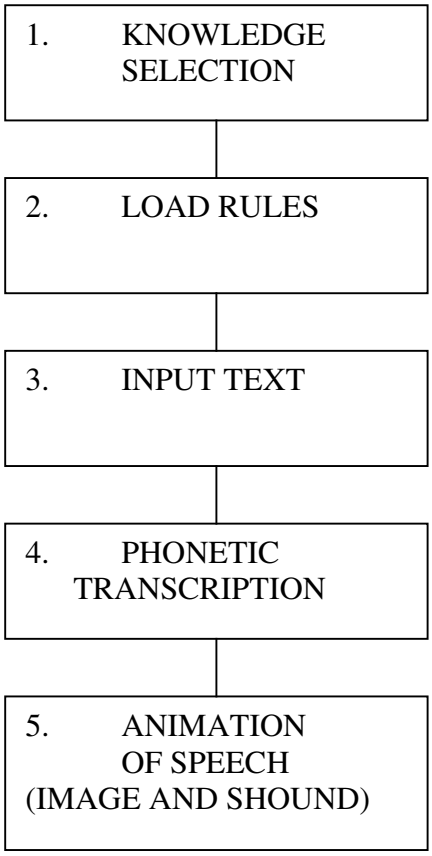


Figure 7. Rule-interpreter: general block diagram

Figure 8. Rule-interpreter's welcoming display screen

TABLE III

Phonetic Transcription of a phrase:

Knowledge : IPASP
Input text: le rogue, Enrique.

Rule Structure

A[B]C		D
Pointer position	Rule used	Phonemes buffered
le rogue, Enrique.	[1]	L
le rogue, Enrique.	[e]	Eh EH1
le rogue, Enrique.	[r]	R1
le rogue, Enrique.	[o]	O
le rogue, Enrique.	[gue]	KV HVC EH EH1
le rogue, Enrique.	[,]	PA
le rogue, Enrique.	[ ]	
le rogue, Enrique.	[e]	EH EH1
le rogue, Enrique.	[nr]	N R1
le rogue, Enrique.	[i]	E E
le rogue, Enrique.	[qu]	K
le rogue, Enrique.	[e]	EH EH1
le rogue, Enrique.	[.]	PA PA

*Each phoneme requires a particular representation of the speech organs. Thus, Babel system has a specific image (of the vocal speech organs) for almost all the 64 Votrax phonemes. See table IV, where the numbers appearing in the columns Front (mouth) and Profile (tongue) are related to Figure 9 and Figure 10 respectively.

TABLE IV

Relationships between Votrax phonemes and Vocal Tract Images

Votrax	Front	Profile	Votrax	Front	Profile
PA	1	1	L	12	12
E	6	10	L1	12	12
E1	10	10	LF	12	12
Y	6	10	W	12	18
YI	6	6	B	3	13
AY	10	10	D	5	12
IE	6	10	K	8	8
I	10	10	P	3	13
A	10	10	T	5	12
A1	9	6	K	4	4

EH	10	10	HV	*	*
EH1	10	10	HVC	8	8
AE	4	2	HF	9	9
AE1	4	2	HFCT(k)	4	4
AH	2	2	HN	*	*
AH1	2	2	Z	17	18
AW	14	9	S	17	18
O	14	9	J	16	16
OU	14	9	SCH	16	16
OO	18	4	V	7	17
IU	18	4	F	7	17
IU1	18	4	THV	17	18
U	18	4	TH	17	18
U1	18	4	M	3	13
UH	2	2	N	5	12
UH1	2	2	NG	11	4
UH2	2	2	:A	*	*
UH3	2	2	:OH	*	*
ER	13	15	:U	*	*
R	13	15	:UH	*	*
R1	13	15	E2	*	*
R2	13	15	LB	*	*

In a few words this program translates text-to-speech by interpreting and applying the letter-to-sound rules (of the knowledge selected) to any input text. Once scanned, the system generates in the synthesizer a smooth bass voice in conjunction with two visual projections of the human face (exhibiting the speech organs), which depict the desirable position of the organs of speech articulation to produce the phonemes determined by the text.

C. Computer Graphics

Two methods were used to create the images for the windows: (1) the vocal tract was drawn graphically, pixel by pixel on the screen using a utility program developed exclusively for this purpose; and (2) several other images were digitized with a Digital VAX 11780 computer and a Gould DeAnza IP 8400 image processor.

Operation and Evaluation

There are two ways to interact with the system:

1.Through the rule-editor to create and update pronunciation rules.

2.Through the rule-interpreter to get speech animation of any input text.

A. Rule-Editor

The creation of the rules is the most important and delicate interaction with the system. In face, the success of the expert system rests entirely on accuracy of the rules. Therefore, it is necessary to invest considerable time with the Rule-editor before satisfactory performance can be achieved.

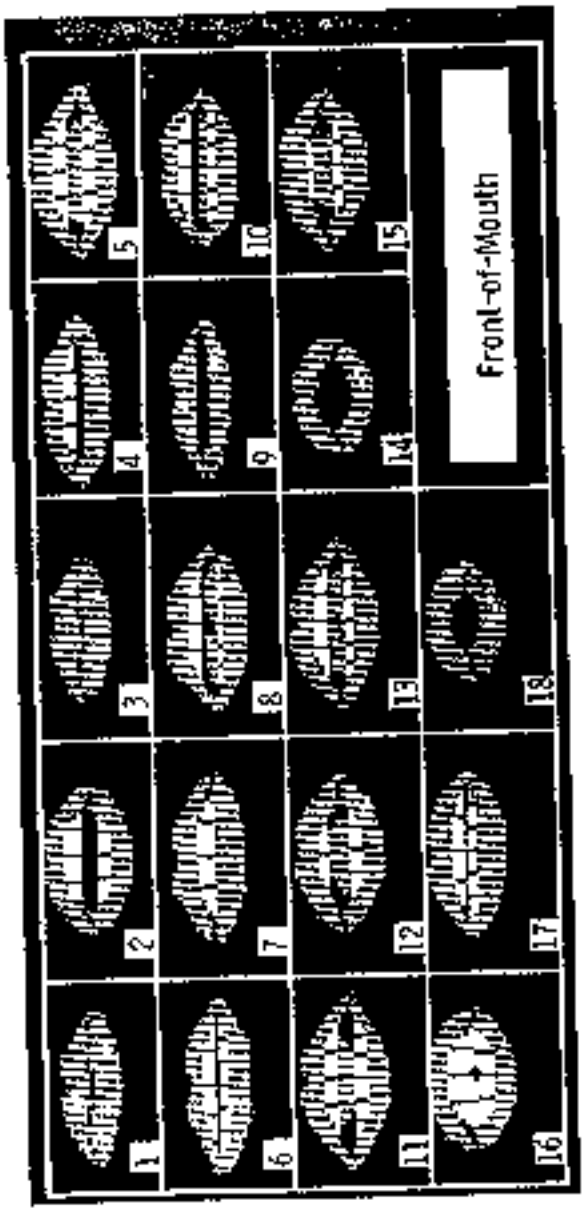


Figure 9. Set of Front-of-Mouth positions available in Babel system.

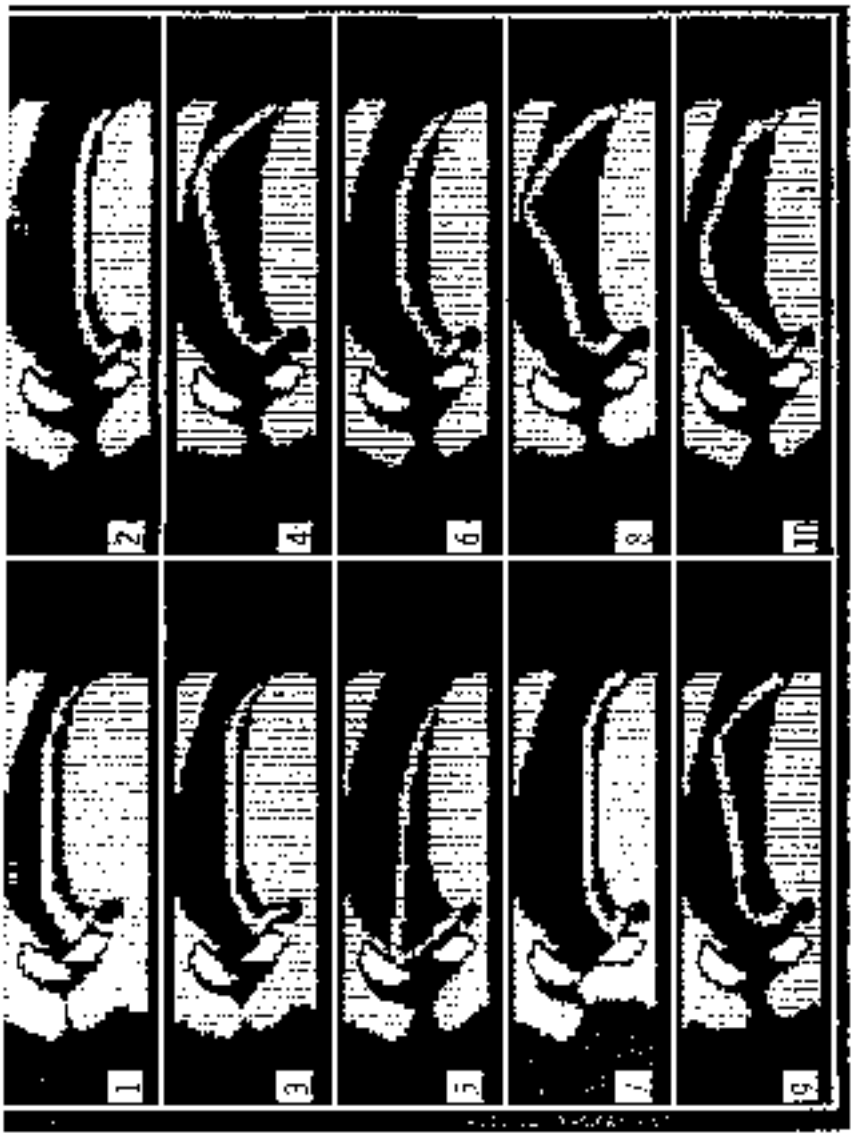


Figure 10. Set of Tongue Positions shown in profile available in Babel system.

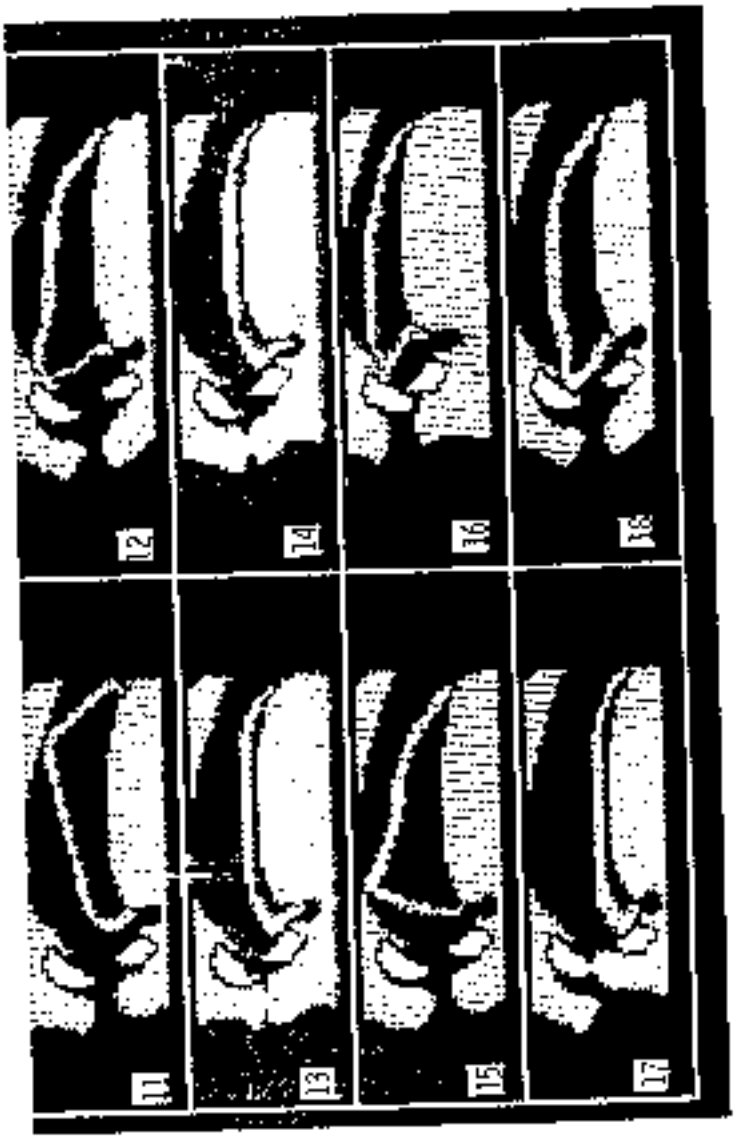


Figure 10. Set of Tongue Positions shown in profile available in Babel system.

* **Rule Development:**

Spanish maintains a fairly good one to one relationship between letters and sounds. Taking advantage of this fact and following the words of Adelstein [1973] and Navarro [1967], the creation of the first draft of rules was feasible. Appendix A is a complete user's manual of the Rule-editor which explains the features of each window as well as how to create and update rules.

The creation of the rules was over as soon as the spoken output of the expert system was understandable and pleasing. However, in several cases due to the limited set of phonemes provided by the synthesizer it was not possible to generate or improve the sounds of some phonemes. For example, the nasal voiced consonants n and ii are currently causing problems in the pronunciation of some words. The phoneme n is provided by the synthesizer but the ii is not. Moreover n is usually confused with the consonant 1. The synthesizer pronounces both n and I as a voiced alveolar, but in human speech an I is a lateral and n is a nasal. It seems that the synthesizer failed to distinguish in its production between sounds with similar points of articulation (but different timbres), especially in allowing the hearer to differentiate between nasal sounds and non-nasal sounds.

At present, there are a set of 68 letter-to-sound rules that translate Spanish text into speech. The name of the knowledge where such rules are preserved is IPASP. The current output of the system can be improved with more exhaustive rules. However, polishing the rules is a task that might take time yet one would be undoubtedly regarded with more pleasing outputs.

B. Rule-Interpreter

The rule-interpreter is designed to animate a human speech model given a knowledge (set of rules) and any input text. The first and only query of the rule-interpreter is the name of the knowledge to be used. Once the knowledge is loaded the user is welcome to type a limited size text that may include any character defined by the rules. The computer will repeat the speech animation as many times as the user keeps pressing any key but <RETURN>. If <RETURN> is pressed, the input window will be erased and the user may type again. When pressing the key Fl after the text has been input, the expert system slows down the animation process in order to let the student appreciate in detail (phoneme by phoneme) the phonetic transcription of the input text. By toggling the Fl key again the expert system returns to its normal animation speed.

Results and Conclusions

Some of the students who have used the Babel system have commented that while interacting with the system, they realized for the first time what was going on inside of their mouth, where their tongue was in the speech process. And they felt it was very easy to use. As stated, some of the applications of Babel might be in the areas of phonetic course training, speech pathology, file-text-

readers, bilingual transcription, showing progressive stages in the process of articulation, and FL instruction. Babel has proved to be a flexible and valuable tool in teaching language pronunciation, offering potential users standardization of knowledge via expert systems.

The fact that the students can see what they hear originates in them an awareness of the speech process. Furthermore, if the students with the guidance of an expert learn to imitate properly the outputs of the Babel system, they will surely undergo a unique learning experience.

References

Adelstein, Miriam. *La Ensenanza del Espanol Como Idioma Extranjero: de la teoria a la practica*. Madrid, Spain: Playor, S.a. 1973. p. 29-81.

Bassnett-McGuire, Susan. *Translation Studies*. New York: Methuen & Co., 1980, p. 13.

Bernstein, J., Pisoni, D.B. "Unlimited Text-to-Speech System: Description and Evaluation of a Microprocessor Based Device." *IEEE-ICASSP*, 1980 p. 576-579.

Bolinger, D.L., Bowen, J.D., Brady, A.M., Haden, E.F., Potson, L., Sacks, N. *Modern Spanish a Project of the Modern Language Association*. New York: Harcourt, Brace and company. 1960, p. 3-4.

Bowen, J.D., Stockwell, R.P. *Patterns of Spanish Pronunciation a Drillbook*. Chicago: The University of Chicago Press. 1960, p.1.

Carlson, R., Granstrom, B., Hunnicutt, S. "A Multi-Language Text-to-Speech Module." *IEEE-ECASSP*, 1982 p. 1604-1607.

Carlson, R., Granstrom, B., Hunnicutt, S. "Bliss Communication with speech or Text Output," *IEEE-ICASSP*, 1982 p. 747-750.

Cater, John P. *Electronically speaking: Computer Speech Generation*. Indianapolis: Howard W. Sams & Co. 1983, p. 74.

Diringer, David. *The Alphabet a key to the history of mankind*. New York: Funk & Wagnalls, 1968, Volume 1, p. 12.

Elovitz, H.S., Johnson, R., McHugh, A. and Shore, J.L. "Letter-to-Sound Rules for Automatic Translation of English Text to Phonetics," *IEEE Transactions on Acoustic Speech and Signal Processing*. December 1976 p. 446-459.

Encyclopedia Britannica. "Phonetics." Chicago: William Benton, 1966, Volume 17 p. 897-900.

Fisher, William M. "Text-to-Speech Development System," *IEEE-ICASSP*, 1983 p. 1344-1347.

Flanagan, James L. "Voices of Men and Machines" *Speech Synthesis* (Reprinted from JASA, 1972, p. 1375). Pennsylvania: Dowden, Hutchinson & Ross, Inc. 2973, p. 9.

Klatt, Dennis H. "The Klattalk Text-to-Speech Conversion System," *IEEE-ICASSP* 1982 p. 1589-1592.

Navarro, Tomas. *Manual de la Pronunciacion Espanola*. New York: Hafner Publishing Company, 1967 p. 13-145.

Navarro, Tomas. *Studies in Spanish Phonology*. Miami: University of Miami Press. 1968, p. 14, 17, 25-26.

Olabe, J.C., Santos, A., Marinez, R., Munoz, E., Martinez, M., Quilis, A., and Bernstein, J. "Real Time Text to Speech Conversion System for Spanish," *IEEE-ICASSP*, 1984 p. 2.101-2.10.3.

Resnick, Melvyn C. *Introduccion a la historia de la lengua espanola*. Washington, D.C. Georgetown University Press. 1981, p. 1.

Santos, J.M., Nombela, J.R. "Text-to-Speech Conversion in Spanish a Complete Rule-Based Synthesis System," *IEEE-ICASSP*, 1982 p. 1593-1596.

Seleskovitch, Danica. *Interpreting for International Conferences*. Washington, D.C. 1978, p. 1.

Steiner, George. *After Babel*. New York: Oxford University Press, 1975, p. xi.

Votalker IB Speech Synthesizer. (A Manual). Votrax, Inc. Artic Technologies: 1985, p. 4-3 to 4-14.

Winston, Patrick Henry. *Artificial Intelligence*. Massachusetts: Addison-Wesley Publishing Company. 1984, p. 164.

Appendix A

Rule-Editor User's Manual

The Rule-editor allows you to create and update letter-to-sound rules. In order to run the Rule-editor or the Parser the Speech Operating System KERNEL must be loaded first (Votalker IB, 1985 p. 2- 1)

The KERNEL is the heart of the Speech Operating System. It provides a software interface between the Botrax SC-02 Chip and application software for tasks like interrupt handling, phoneme buffering, data format expansion and lexicon management. To load the Speech Operating System

KERNEL simply type...

SONIX <return>

In most cases an autoexec.bat file will load the Speech Operating System KERNEL. Nonetheless, if KERNEL is not loaded and you attempt to run the Rule-editor or the Parser the following message will appear.

CANNOT LOCATE THE SONIX KERNEL!
RETURN TO DOS AND LOAD SONIX

A. Create and Edit A Knowledge

1.Run EDITOR (Rule-editor) by typing: EDITOR <return>

2.The Main menu will appear. See Fig. 3. The Main menu presents a set of options to create and edit knowledge as well as to define and retrieve UDS (user defined symbols).

If you press:
-Q key (QUIT): The session is ended.
-L key (LOAD): A prompt will appear in the INPUT/OUTPUT AREA at the bottom of the screen asking for the Knowledge to load. -S key (SAVE): A prompt will appear in the INPUT/OUTPUT AREA at the bottom of the screen asking for the Knowledge to save.
-Fl key (UDS N-OF): A prompt will appear in the INPUT/OUTPUT AREA at the bottom of the screen. See Fig 7. It requests first any permitted UDS symbol; next a limited-range integer number; and eventually a set of character strings that will define such a symbol. -F2 key (UDS N-OR-MORE): A prompt will appear in the INPUT/OUTPUT AREA at the bottom of the screen. See Fig. 7. It requests first any permitted UDS symbol; next a limited-range integer number; and eventually

a set of character strings that will define such a symbol.
-F6 key (DISPLAY UDS): A column will be displayed in the INFORMATION AREA with all the permitted UDS symbols that can be selected using the selecting arrow (this is manipulated with the up and down arrows of the keyboard). Once the target symbol to display is selected just hit <return>. See Fig. 7.
-E key (EDIT): To get in the Edit mode. The Edit menu will appear. See Fig. 4. At the top of the WORKING RULE AREA you can identify the A,B and C parts of the rules. That means the prefix, body and suffix rule respectively. The remaining part of the WORKING RULE AREA exhibits the D component of the rule structure, which is a set of phonemes denoting the sounds of the rule. The Botrax phonemes can be altered by means of eight parameters. In other words, the parameter manipulation allows the generation of phoneme variations called allophones. Fig. I I shows the meaning of the parameters (abbreviation at the top of the WORKING RULE AREA) and their range of values.

The Edit Menu offers a set of commands to create and edit rules.
If you press:
*Esc key (MAIN), you return to the Main menu.
*Ins (INSERT), you will split the current set of phonemes to insert a new one just after the cursor location.
*Del (DELETE), the phoneme where the cursor is positioned will be deleted.
*+ (NEXT), the next rule (if any) will be displayed. *- (PRIOR), the previous rule will be displayed. *Ctrl T (Edit title-rule) you will be free to develop a rule of the form A[B]C. Remember you have to enclose B (the body rule) in brackets. As mentioned, if A and C are omitted, it means that the context is irrelevant.
*F7 (NUMBER), you can get any rule in the WORKING RULE AREA by typing after a prompt that will exhibited in the INPUT/OUTPUT AREA. If you choose any rule-number that does not exist a warning message will be displayed.
*F 10 (SPEAK), you will listen to the set of phonemes appearing in the WORKING RULE AREA.
*F3 (PHONEME CHART), the 64 Votrax phonemes will be displayed in the INFORMATION AREA. Each phoneme is labeled with a character string. See Fig. 4.
*F4 (CHARACTERS), all the special letters and accents available will be displayed in the INFORMATION AREA . This chart indicates how to get such characters if they are needed when defining a rule (A[BIC). See Fig. 5 and Fig. 6.

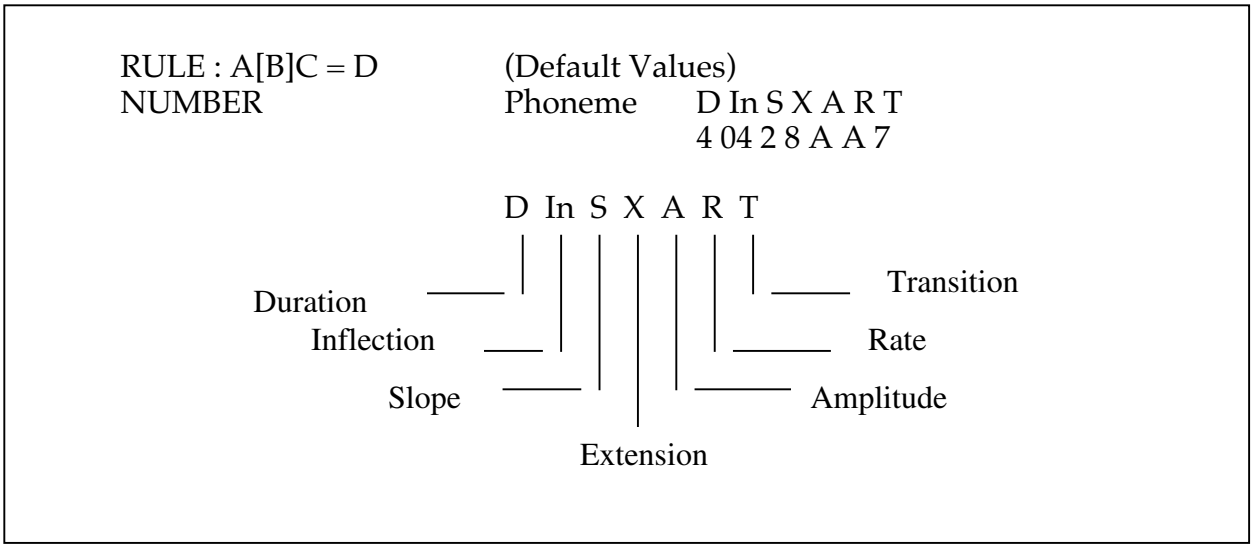
*Alt Fl (INS RULE), will insert a dummy rule between the current rule and the previous rule. This option helps in having a particular sequence of the rules, because the order of the rules is significant.
*Alt F2 (DEL RULE), erases a rule from the knowledge.

"Rule: A[B]C=D (Default Values)"
Number
Phoneme D IN S X A R T
4 04 2 8 A A 7

D IN S X A R T
Duration_____ Transition

Inflection Rate
Slope Amplitude
Extension

Duration: This parameter allows 4 level of phoneme duration. Acceptable rank: I to 4. (maximum to minimum, 4 is used for fullest phoneme pronunciation).
Inflection: Allows variation in the pitch (tone) of the phoneme. Acceptable rank: 00 to IF (low to high, 32 possible inflection levels).
Slope: "The Slope of the Inflections is the parameter that determines the rate at which inflections glide from one level to another" (Votalker IB, 1985, p. 4-1 1). Acceptable rank: 0 to 7.
Extension: "With Pitch Extension you can raise or lower the pitch in smaller intervals than inflection levels allow..." (Votalker IB, 1985, p. 4-14). Acceptable rank: 0 to F.
Amplitude: This parameter allows to stress or accent the phonemes. Acceptable rank: 0 To F (quiet to loud, 16 amplitude levels).
Rate: "Rate adjustments are made to accommodate the fine tuning of a phoneme sequence's pronunciation" (Votalker IB, 1985, p. 4-12). Acceptable rank: 0 to F (slow to fast, 16 rate levels).
Transition: "The Formant Transition controls the speed at which bands of frequencies in the synthesizer move." (Votalker IB, 1985, p. 4-14).



Duration: This parameter allows 4 level of phoneme duration. Acceptable rank: I to 4. (maximum to minimum, 4 is used for fullest phoneme pronunciation).

Inflection: Allows variation in the pitch (tone) of the phoneme. Acceptable rank: 00 to 1F (low to high, 32 possible inflection levels).

Slope: "The Slope of the Inflections is the parameter that determines the rate at which inflections glide from one level to another" (Votalker IB [1985], p. 4-11). Acceptable rank: 0 to 7.

Extension: "With Pitch Extension you can raise or lower the pitch in smaller intervals than inflection levels allow..." (Votalkel IB [1985], p. 4-14). Acceptable rank: 0 to F.

Amplitude: This parameter allows to stress or accent the phonemes. Acceptable rank: 0 to F (quiet to loud, 16 amplitude levels)

Rate: "Rate adjustments are made to accommodate the fine tuning of a phoneme sequence's pronunciation" (Votalker IB, p. 4-12). Acceptable rank: 0 to F (slow to fast, 16 rate levels).

Transition: "The Formant Transition controls the speed at which bands of frequencies in the synthesizer move-" (Votalker IB, p. 4-14).

Figure 11. Votrax Phoneme Parameters

Authors' Biodata

Joaquin Vila is an Assistant Professor of the Applied Computer science Department at Illinois State University. He did his undergraduate work in computer science at Monterrey Tech (Mexico) and his M.S. in computer science and Ph.D. in Engineering Management at the University of Missouri—Rolla. He has written several articles in the computer science field. His research interests are in the areas of artificial intelligence, education, and computer-integrated manufacturing.

Lon Pearson (Ph.D., UCLA, 1973) is a Professor of Spanish at the University of Missouri—Rolla. He began working with computers at Johns Hopkins University during a postdoctoral in 1975 and with microcomputers at BYU during a sabbatical in 1982. He has published numerous articles and a book on Chilean literature.

Authors' Addresses

Joaquin Vila
Department of Applied Computer Science
Stevenson Hall 133
Illinois State University
Normal, IL 61761
(309) 438-8338

Lon Pearson
Department of Arts and Cultural Studies
University of Missouri—Rolla
Rolla, MO 65401
(314) 341-4623